

What can Ontology Do for Public Health?

György Surján



Semantic Interoperability and Data Mining in Biomedicine

EU 6FW Program IST 507505

Summer-school and workshop

Balatonfüred, July 3-10.

2004

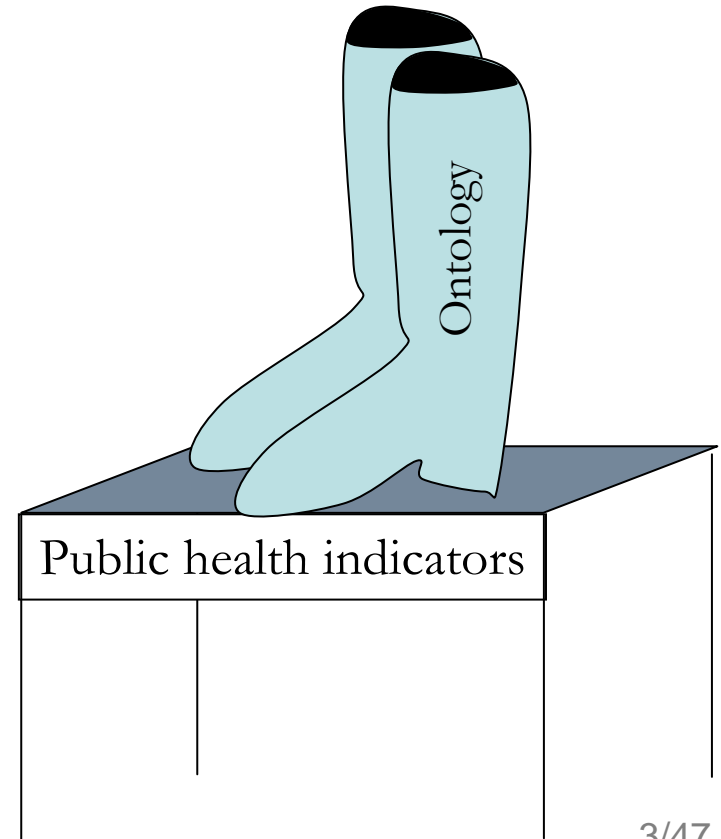
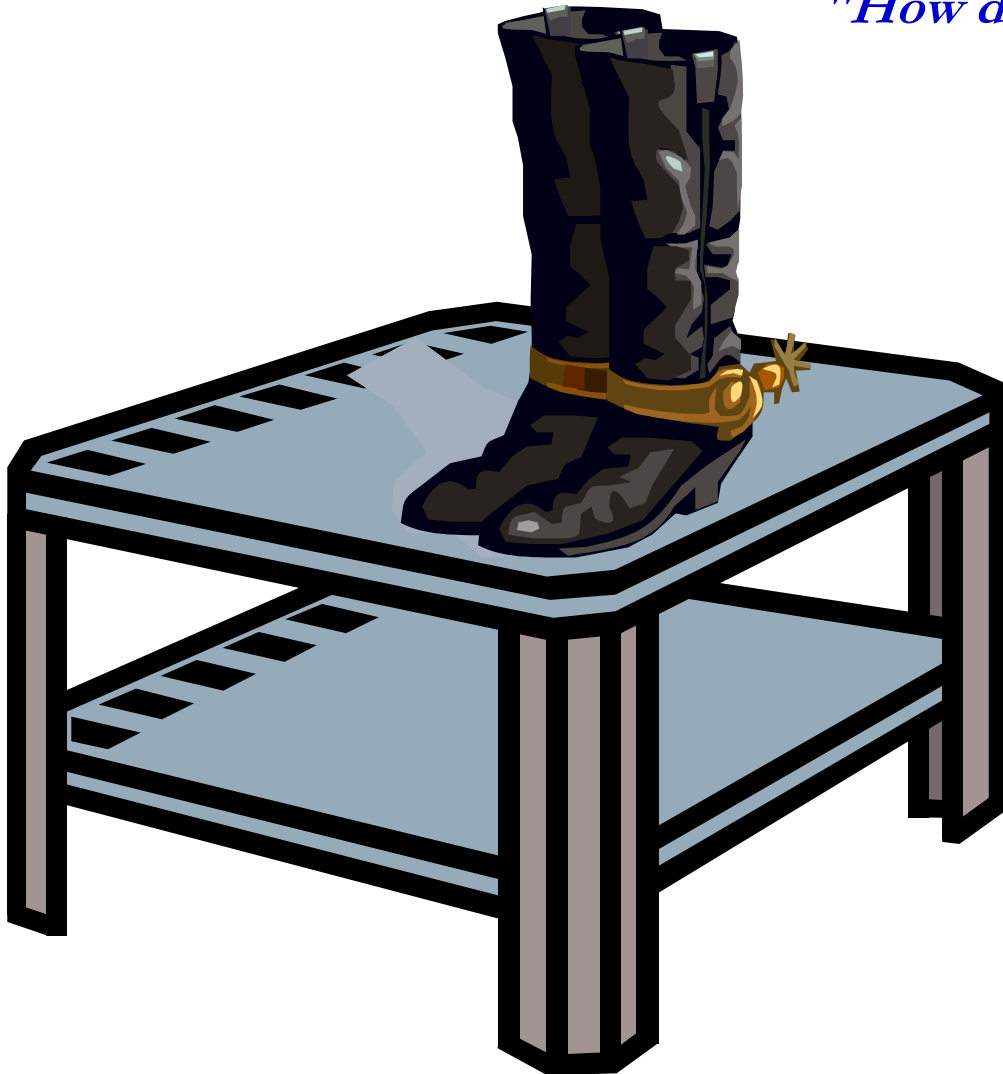
Overview

1. The problem in general
2. Populations
3. Indicators: entities and dimensions
4. Time
5. Categorisations
6. Conclusions

1. The problem in general

What can ontology do for public health?

"How does the boot get on the table?"



Public health needs information

**We do not speak about public health in general
but about public health information.**

**Data stored in public health databases often called
(health) indicators**

Why?

I-1. Why EC health indicators?

The European Commission's Health Monitoring Programme (HMP) was established in 1997 to take forward the enhanced public health responsibilities of the EU in the public health field. It has as its objective *'to contribute to the establishment of a Community health monitoring system'*, in order to:

1. **Measure health status**, its determinants and the trends therein throughout the Community;
2. **Facilitate the planning**, monitoring and evaluation of Community Programmes and actions; and
3. Provide Member States with appropriate health information to **make comparisons** and support their national health policies.

What?

ECHI does not provide definition

For our purpose:

*A public health indicator is any data or information which characterises health condition of a **population** or factors that influencing it.*



A public health indicator is any data or information which characterises health condition of a population or factor that influencing it.

What is the problem?

“3. Provide Member States with appropriate health information to **make comparisons** and support their national health policies.”

Need of comparability

“You cannot compare apples with peaches”





What should be compared?

- **Two different populations (in the same time)**
- **The same population in different time**

1. The problem in general
- 2. Populations**
3. Indicators: entities and dimensions
4. Time
5. Categorisations
6. Conclusions

Comparison of two different populations

Are Hungarian people more healthy than Estonian?

What does it mean: Population of Hungary ?

- Hungarian citizens
- People living in Hungary
- People actually staying in Hungary
- All Hungarians?
- Sum of the population of all Hungarian counties
Homeless persons? People with unknown home address?

Comparison of the same population in different time

Are Hungarian people more healthy in 2003 than in they were in 1993?

People in Hungary in 1993 and in 2003 are the same population?
During 10 years many of them died, many were born, all others aged 10 years

The problem of self-identity of a population
(*“Is Mr X more healthy in 2003 than he was in 1993?” – “Is he still the same person?” - Many of his cells died, many new appeared*)

We have to know, what a certain population is.
The science deals with the problem what things are, called ontology.



The whole problem of population will come back later again.

1. The problem in general
2. Populations
- 3. Indicators: entities and dimensions**
4. Time
5. Categorisations
6. Conclusions

E.G.

INCIDENCE: is an observation of a case of a disease not known before.

Incidence as an indicator is the sum of such observations in a given population and in a given period of time.

⇒ **CANCER INCIDENCE** is the number of observed new cancer cases

Is such a definition sufficient?

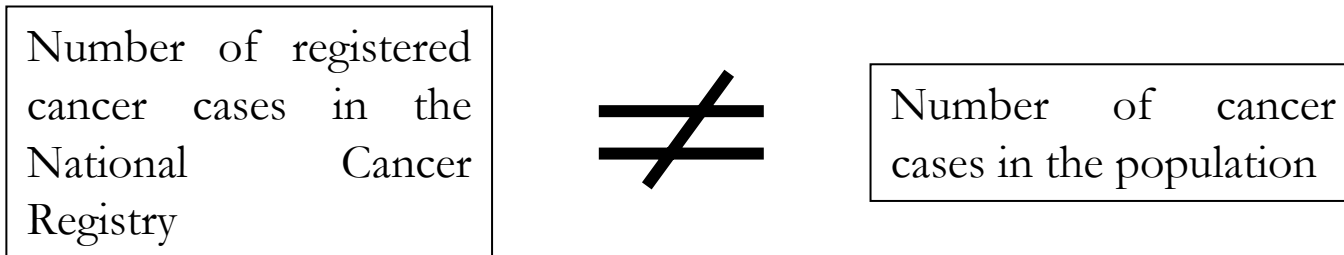
We have to know what a cancer is \Rightarrow need for *medical* ontology

BUT!

We also have to know which cases had been recorded in a registry

\Rightarrow

need for *ontology of public health data*



What is “cancer incidence”?

Recommendations for coding Incidence Date

Distributed in 1995, revised in 1997

The date of the first event (of the six listed below) to occur chronologically should be chosen as incidence date. If an event of higher priority occurs within three months of the date initially chosen, the date of the higher priority event should take precedence.

Order of declining priority:

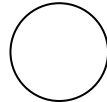
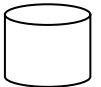
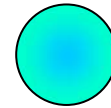
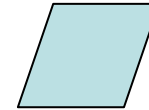
1. **Date of first histological or cytological confirmation** of this malignancy (with the exception of histology or cytology at autopsy). This date should be, in the following order:
 - a) date when the specimen was taken (biopsy)
 - b) date of receipt by the pathologist
 - c) date of the pathology report.
 2. **Date of admission to the hospital** because of this malignancy.
 3. When evaluated at an outpatient clinic only: **date of first consultation** at the outpatient clinic because of this malignancy.
 4. **Date of diagnosis**, other than 1, 2 or 3.
 5. **Date of death**, if no information is available other than the fact that the patient has died because of a malignancy.
 6. Date of death, if the malignancy is discovered at autopsy.
- Whichever date is selected, the date of incidence should not be later than the date of the start of the treatment, or decision not to treat, or date of death.
- The choice of the date of incidence does not determine the coding of the item "basis of diagnosis".

The problem is not the ontology of time,
but the ontology of data



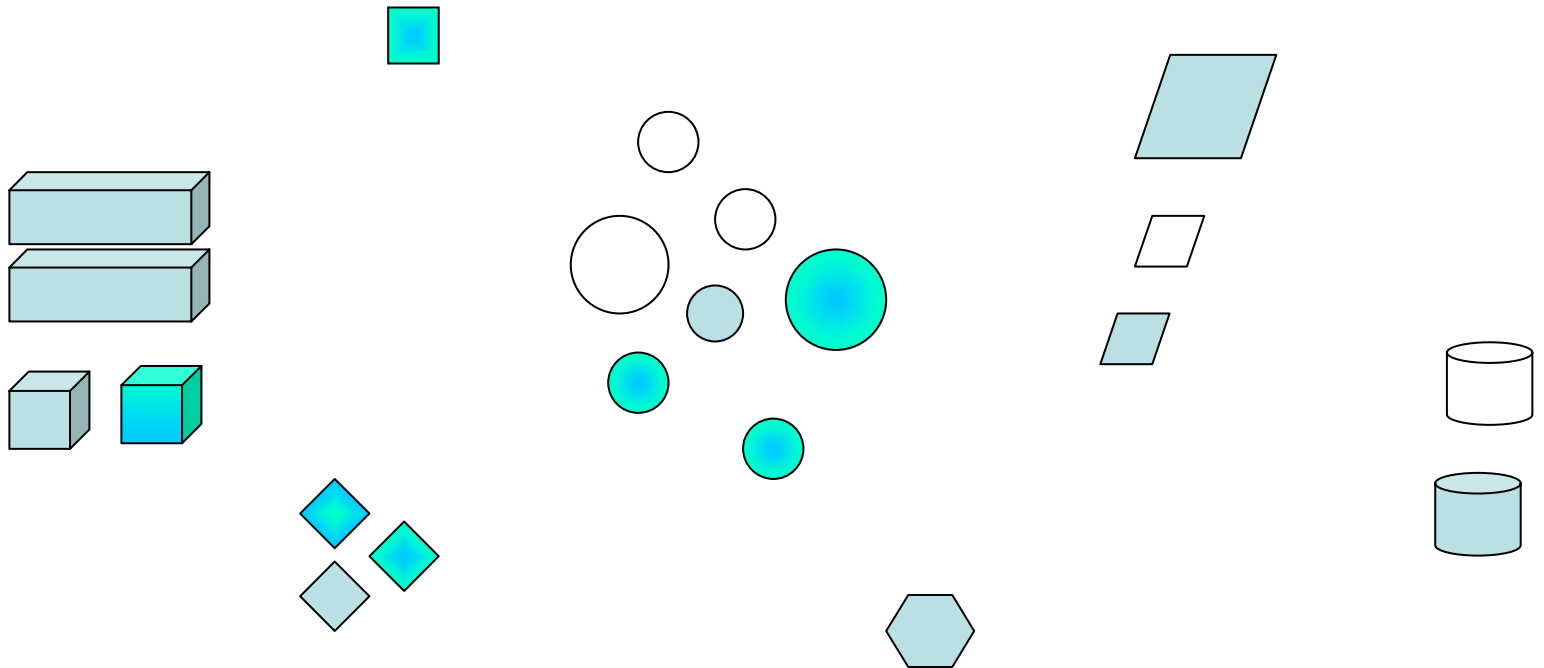
Things that are observed
and enumerated

Things have various properties



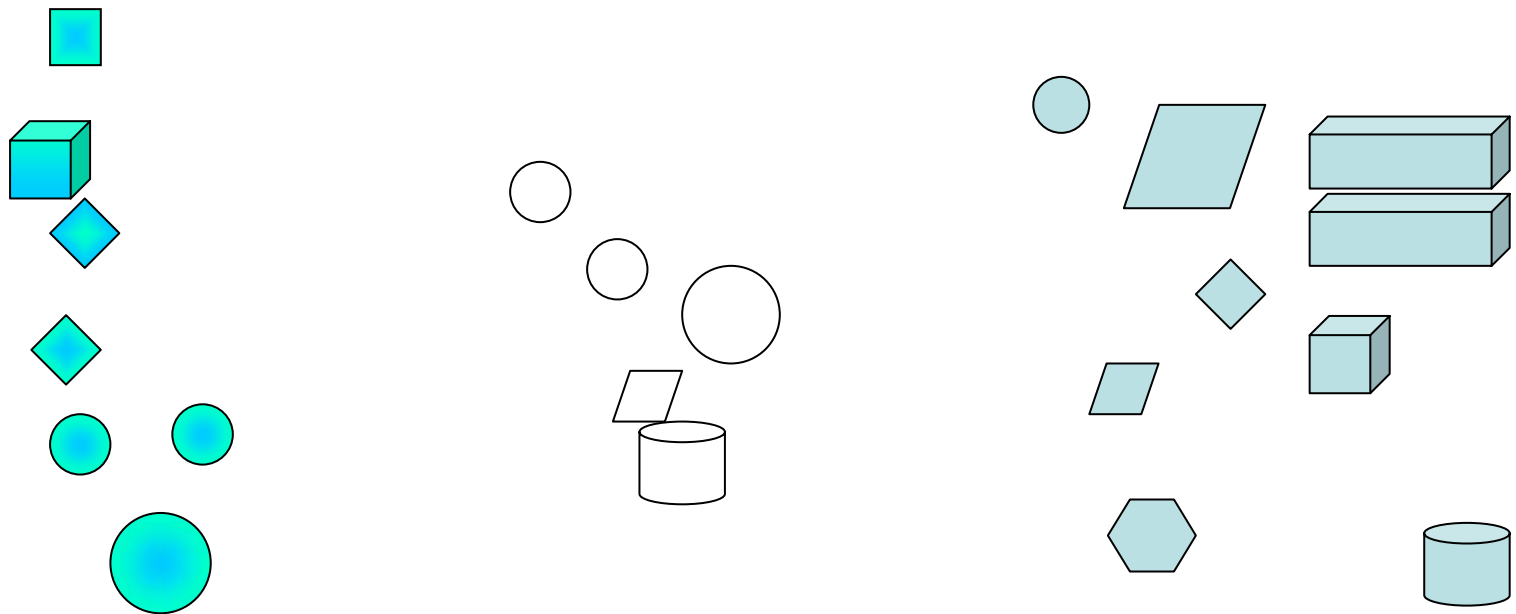
What are the “things” and what are their “properties”
“substance” and “attribute” or categorisations

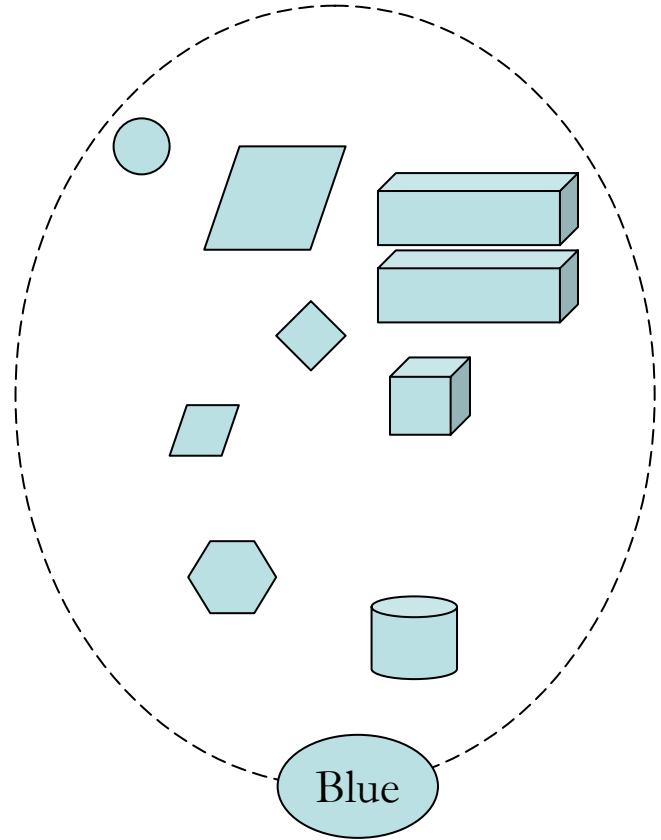
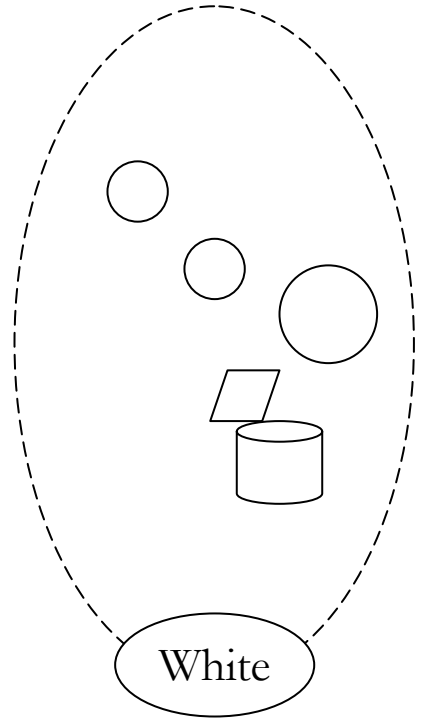
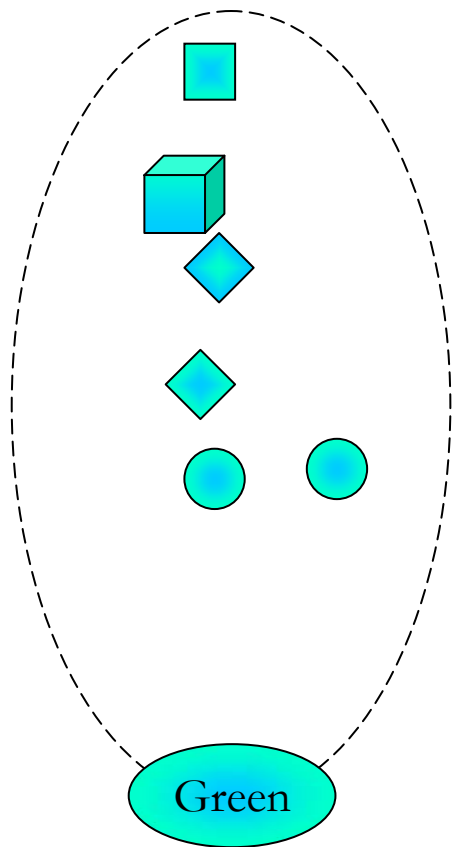
“things” = forms, “properties” = colour, size



What are the “things” and what are their “properties”
“substance” and “attribute” or categorisations

“things” = colours, “properties” = shape, size





A Cylinder WhichHasColor Blue

A Blue thing WhichHasShape Cylinder



A Cylinder WhichHasColor Blue

A Blue thing WhichHasShape Cylinder

“All these are called substances because they are not predicated of a subject but everything else is predicated of them”

Aristotle
Metaphysics, Book V. 8.

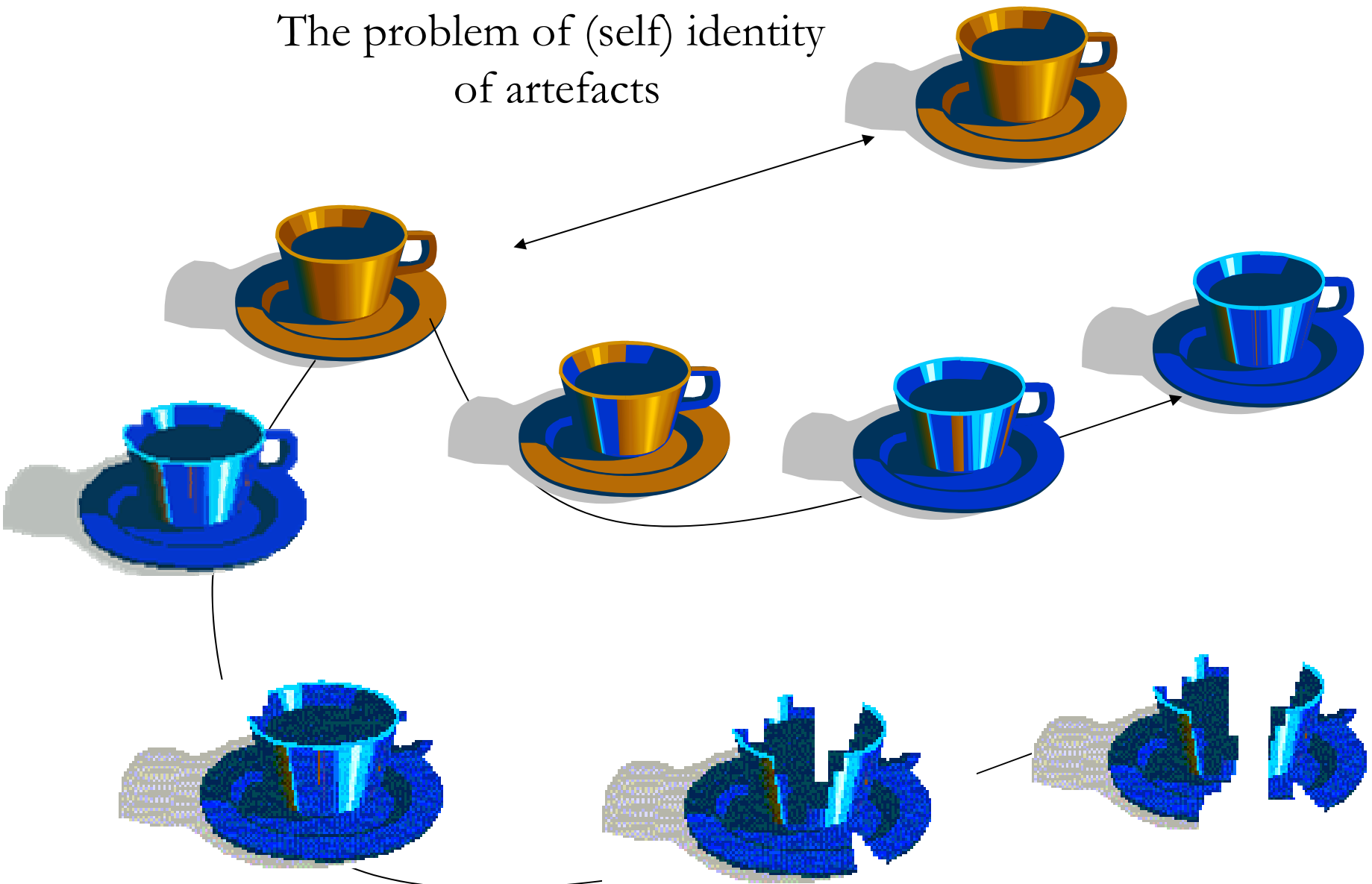
The problem of (self) identity



The continuous feeling of self identity

**We are able to imagine ourselves in different roles,
but hardly can imagine ourselves not being ourselves**

The problem of (self) identity of artefacts



The problem of (self) identity of public health indicators

Public health indicators are artefacts



A Cylinder WhichHasColor Blue

two different "substances"
having different predicates

~~A Blue thing WhichHasShape Cylinder~~

traffic accident

Death

statistics on
traffic
accidents

lethal
outcome



of traffic
accident origin

mortality
statistics

space?
time ?



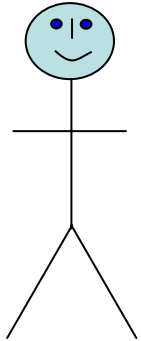
Entities that are observed necessarily exist in space and time

Space and Time are basic dimensions

All other dimensions depend on the nature of the observed thing

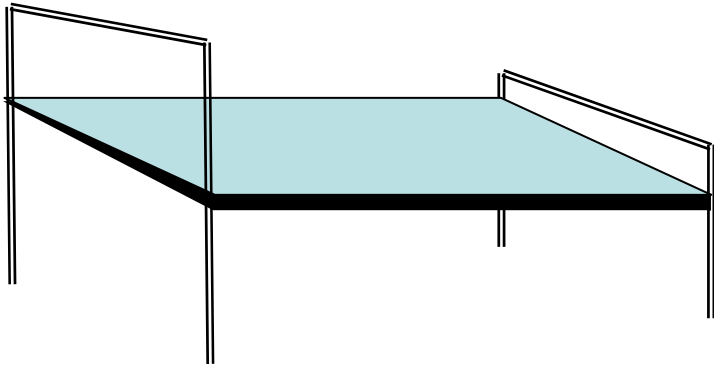
OBSERVED ENTITIES

DIMENSIONS



Person

Sex
Age (birth date)
Home address
Occupation
Etc.



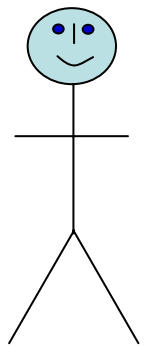
Hospital bed

Technical parameters (wheels, fixed legs etc.)
Medical specialty.

OBSERVED ENTITIES

DIMENSIONS

necessary attributes
cannot change
without loss of identity



Person

contingent attributes
may change
without loss of identity

Sex
Age (birth date)

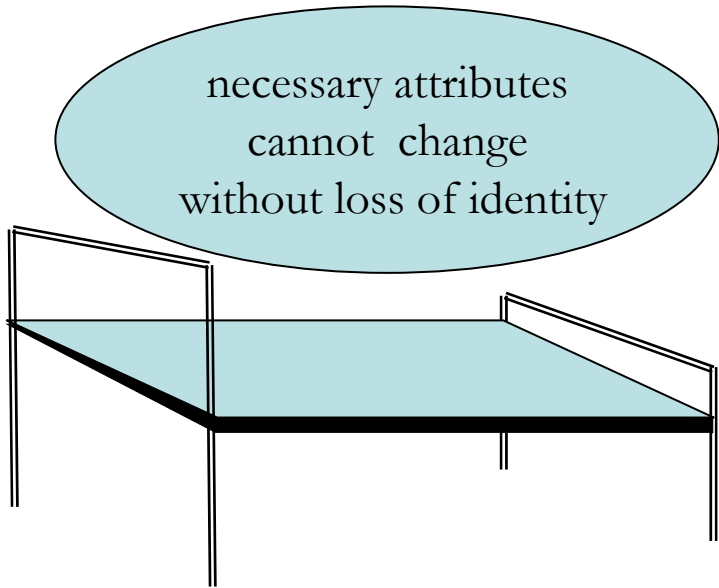
Home address
Occupation
Etc.

IDENTITY

**AT THE TIME OF
THE OBSERVATION**

OBSERVED ENTITIES

DIMENSIONS



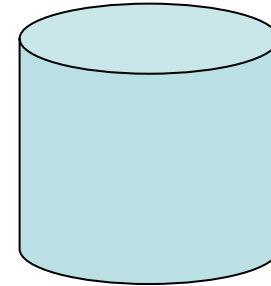
necessary attributes
cannot change
without loss of identity

Hospital bed

contingent attributes
may change
without loss of identity



Technical parameters
(wheels, fixed legs etc.)
Medical specialty.



Identity of the observed entities.

Identity of the indicators

Ontology of the public health indicators

What belongs to the “indicator” and what to the dimensions?

Cancer incidence (age, sex, location,..)

Incidence (disease, age, sex, location,..)

↳ {cancer}

Identity 1.

**Two indicators are identical if the observed entities
and all dimensions are the same.**

Two things are identical if they are the same.

Identity 2. *To be pragmatic*

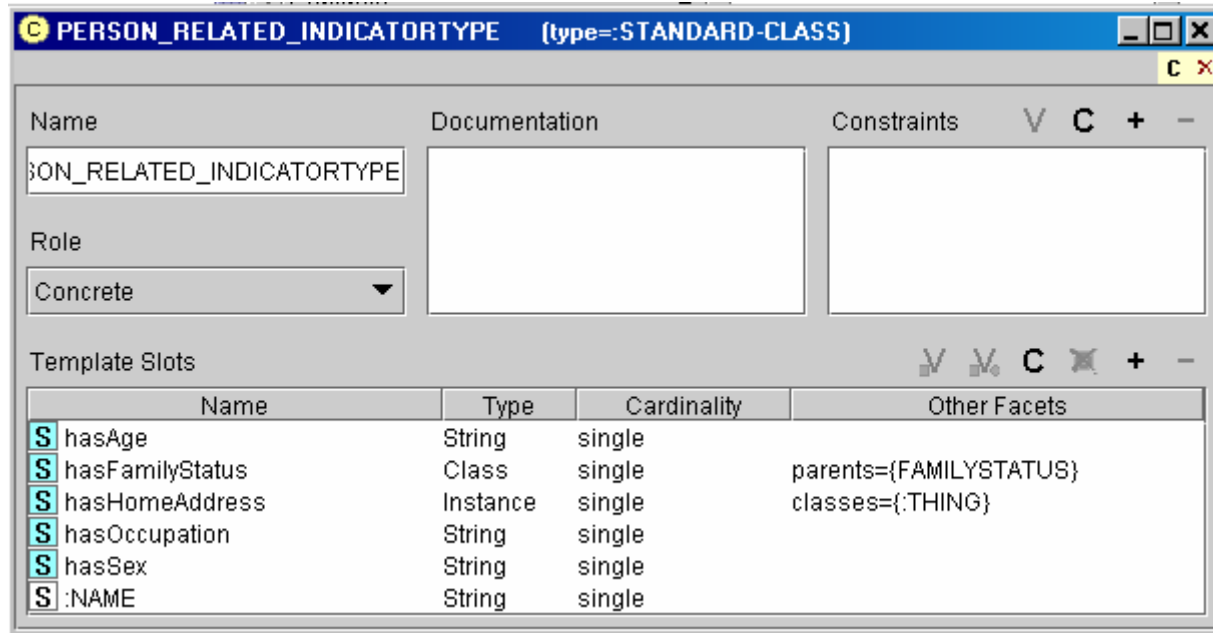
Two “indicators” are identical if they characterise the population from the same aspect.

Number of heavy smokers $><$ smoked μg nicotine /person



neither identical, nor comparable

Ontology of indicators vs. ontology of "things"



We do not have to know (everything about) what *person* is, but we have to know everything about indicators which refer to persons

What does it mean: Population of Hungary ?

- Hungarian citizens
- People living in Hungary
- People actually staying in Hungary
- All Hungarians?
- Sum of the population of all Hungarian counties

Homeless persons? People with unknown home address?

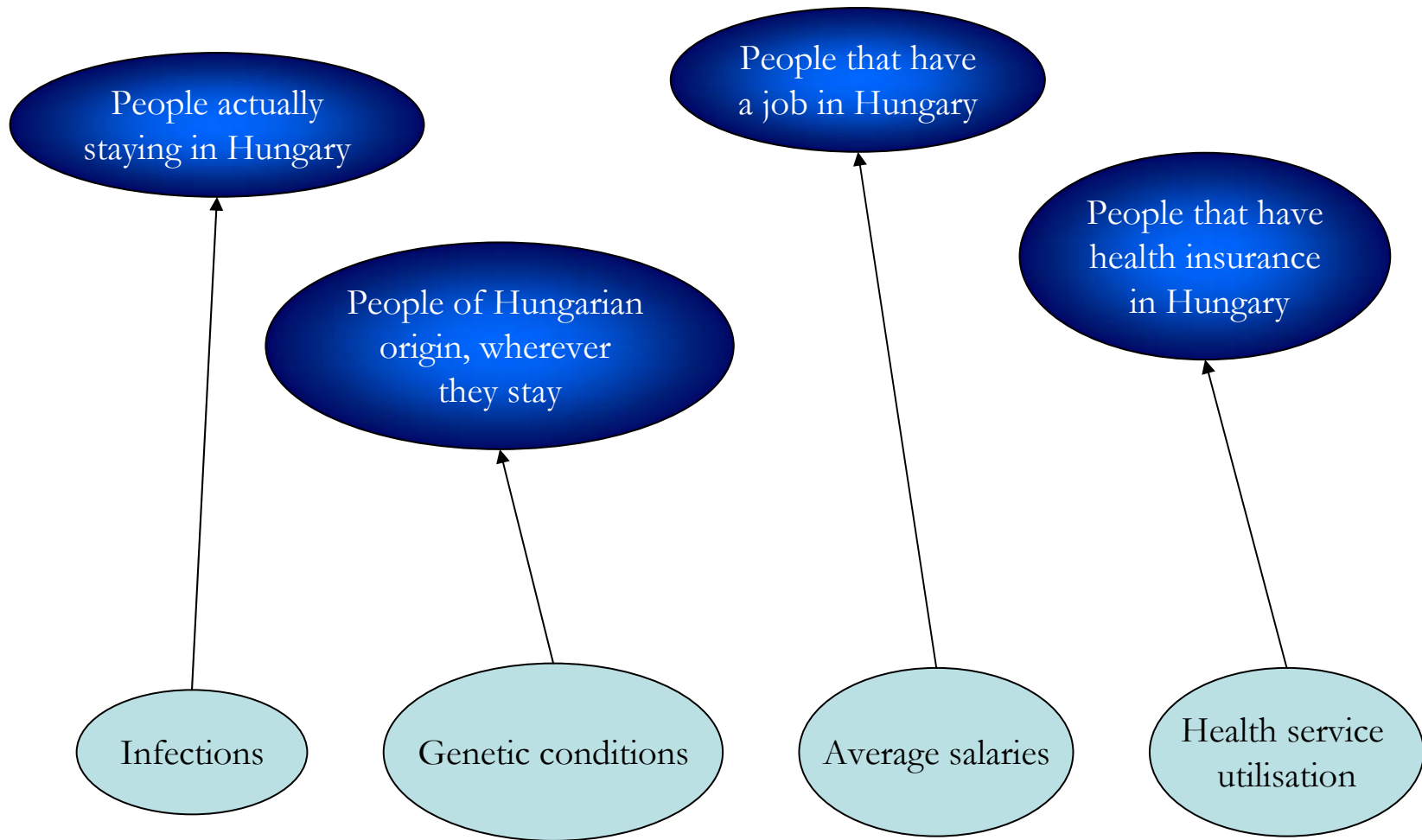
Infections

Genetic conditions

Average salaries

Health service
utilisation

**What does it mean: Population of Hungary ?
Ontological knowledge is not sufficient**



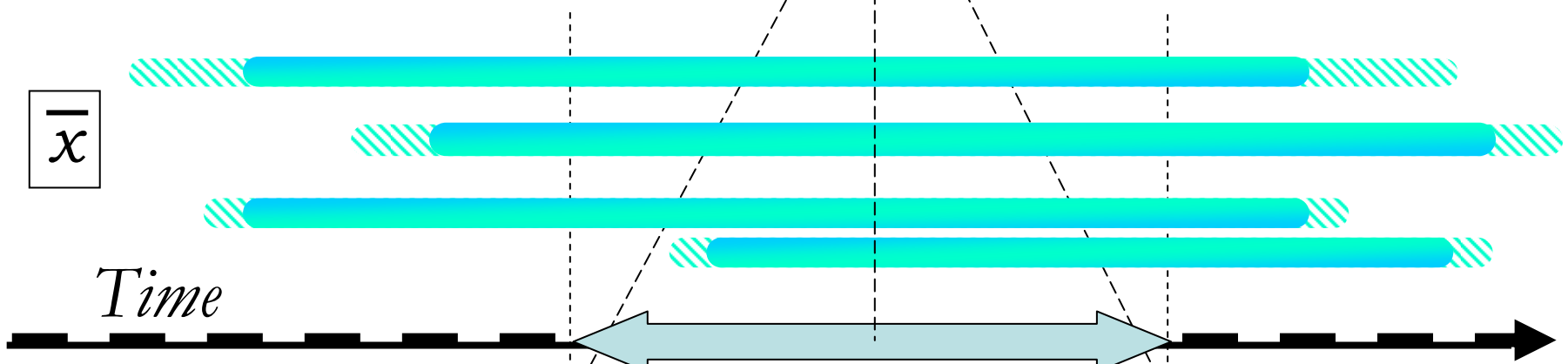
1. The problem in general
2. Populations
3. Indicators: entities and dimensions
- 4. Time**
5. Categorisations
6. Conclusions

Time point, time period,
observation time



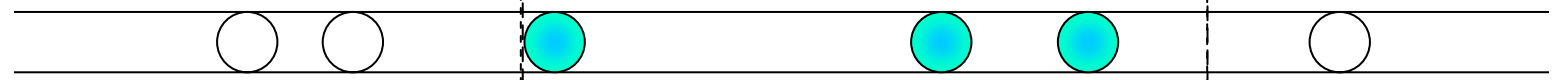
Things

Continuants



\bar{x}

Time



Σx

Occurrants

1. The problem in general
2. Populations
3. Indicators: entities and dimensions
4. Time
- 5. Categorisations**
6. Conclusions

The problem of categorisations 1.

World Health Organisation HFA Health for All database

Indicators

- + 01 DEMOGRAPHIC AND SOCIO-ECONOMIC INDICATORS
- + 02 MORTALITY-BASED INDICATORS
- + 03 MORBIDITY, DISABILITY AND HOSPITAL DISCHARGES
- + 04 LIFE STYLES
- + 05 ENVIRONMENT
- + 06 HEALTH CARE RESOURCES
- + 07 HEALTH CARE UTILIZATION AND EXPENDITURE
- + 08 MATERNAL AND CHILD HEALTH
- ... 09 Unused group
- ... 10 Unused group
- + 11 Temporary indicators

Disjoint categories ?

The problem of categorisations 2.

Health Data 2004

File ?

Selection of chapter

Chapter	Variable	Unit
?	+	-
+ Health Status		
+ Health Care Resources		
+ Health Care Utilisation		
+ Expenditure on Health		
+ Health Care Financing		
+ Social Protection		
+ Pharmaceutical Market		
+ Non-Medical Determinants of Health		
+ Demographic References		
+ Economic References		

Selection of variable

Chapter	Variable	Unit
? Help on the icons		
	Gross domestic product	
	Government final consumption expenditure	
	Final consumption expenditure of households	
	Gross fixed capital formation	
	General government total outlays	
	Public revenue	
	Compensation of employees	
	Average earnings of production worker	
	Government budget outlays for total R&D	

Disjoint categories ?

The problem of categorisations 3.

The screenshot shows a software interface titled "Health Data 2004". At the top, there is a menu bar with "File ?". Below it is a section titled "Selection of chapter" with three tabs: "Chapter", "Variable", and "Unit". Under the "Chapter" tab, there are three buttons: "?", "+", and "-". A tree view of categories is displayed below, with "Tobacco consumption" highlighted in blue. The categories are:

- Health Status
- Health Care Resources
- Health Care Utilisation
- Expenditure on Health
- Health Care Financing
- Social Protection
- Pharmaceutical Market
- Non-Medical Determinants of Health
 - Life styles and behaviour
 - Food consumption
 - Alcohol consumption
 - Tobacco consumption**
 - Body weight and composition
 - Environment: air quality
- Demographic References
- Economic References

Functional vs. ontological categorisation

Tobacco consumption IsA Non-medical determinant of health ?



Let people find their way

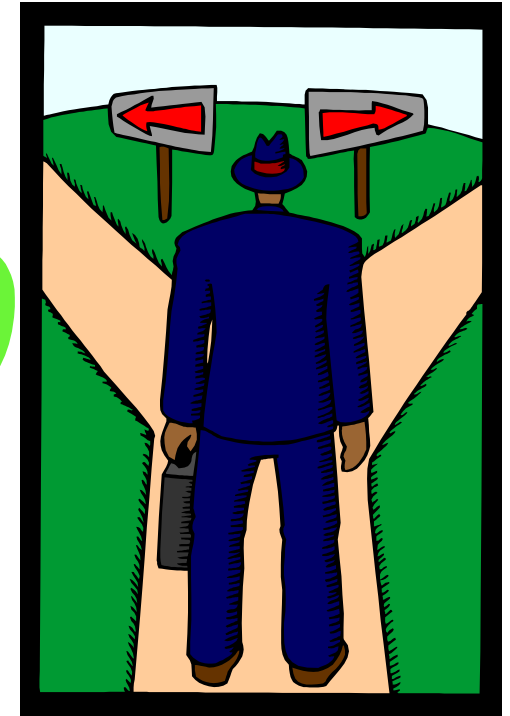
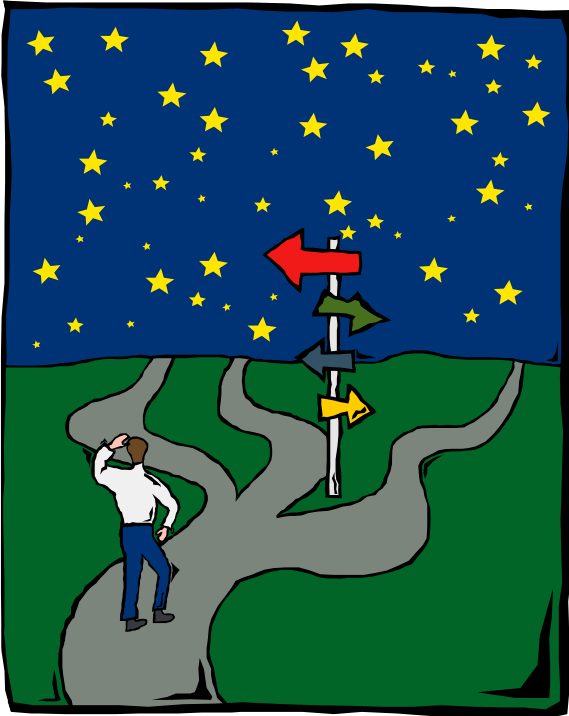
“All routes lead to Rome”

or at least there is more than one route to Rome

What things are and how things can be found are different issues

Functional vs. ontological categorisation

Let people find their way



Various people have various need and preferences

1. The problem in general
2. Populations
3. Indicators: entities and dimensions
4. Time
5. Categorisations
- 6. Conclusions**

Conclusions

1. Lack of ontological background might lead to loss of consistency and comparability in public health databases
2. Ontology of the “things” about data are collected, and ontology of data (indicators) are also important
3. Ontological knowledge alone is not sufficient to create comparable public health databases
4. Ways to find things and structuring things are different issues